Speaker: Jiantao Jiao  
*Department of Electrical Engineering, Stanford University*

Title: From Maximum Likelihood to Minimax Estimation: Enlarging the Effective Sample Size from $n$ to $n \ln n$

Abstract:

We propose a general approach to estimating functionals of discrete distributions, and illustrate its efficacy by analyzing the cases of estimating the entropy $H(P) = \sum_{i \in S} -p_i \ln p_i$ and $F_\alpha(P) = \sum_{i \in S} p_i^\alpha$, $\alpha > 0$, based on $n$ i.i.d. samples from the unknown discrete distribution $P$ over the alphabet $S$ of size $|S|$. The resulting estimators achieve the minimax $L_2$ rates in all regimes, including those where the alphabet size $|S|$ grows with $n$. They have low complexity and do not assume knowledge of $S$ or $|S|$. The performance of these estimators with $n$ samples turns out to be essentially that of the respective Maximum Likelihood Estimators (MLE) with $n \ln n$ samples. Approximation theory plays key roles in analyzing the MLE (the plug-in rule for functional estimation), and in both the construction and analysis of the minimax rate-optimal estimators. As corollaries, we show how our results are consistent with other recent results, such as the optimal sample complexity $n = \Theta(|S|/ \ln |S|)$ for entropy estimation, first discovered by Valiant and Valiant in 2011.

We further demonstrate that our approach can seamlessly accommodate other functionals of interest, such as mutual information. We discuss various algorithms in machine learning, image ranking, medical imaging, computer vision, systems biology, and neuroscience, which naturally involve - either explicitly or implicitly - mutual information estimation, where the current prevailing approach is to use the empirical mutual information. Notable examples include the Chow–Liu algorithm and the Tree-Augmented Naive Bayes (TAN) classifier. Experiments with these and other algorithms show that replacing the empirical mutual information by the proposed estimator results in consistent and substantial performance boosts on a wide variety of datasets.

Based on joint work with Kartik Venkat (Stanford EE), Yanjun Han (Tsinghua University), and Tsachy Weissman (Stanford EE).